ABSTRACT

A computer implemented software method enables the prediction of the pK_a of an arbitrary molecule based upon a knowledge of the molecular structure of that molecule and a statistical analysis of the molecular structures of a group of molecules (training set) for which the pK_a is known. Hierarchical atom connectivity trees are constructed for the training set and the various atoms types identified in each molecule are associated in a bit string for that molecule and also associated with the experimentally determined pK_a for that molecule. PLS analysis of the training set data yields coefficients associated with each atom type represented in the bit strings. A hierarchical atom connectivity tree may then be constructed for the molecule of interest. The predicted pK_a is determined by multiplying the number of occurrences of each atom type in the molecule of interest by the PLS coefficient determined for that atom type and summing the resulting multiplications.